Laughter Type Recognition from Whole Body Motion

Laughter Type Recognition from Whole Body Motion

Harry J. Griffin*, Min S. H. Aung*, Bernardino Romera-Paredes*, Ciaran McLoughlin*
Gary McKeown†, William Curran†, Nadia Bianchi-Berthouze*
*UCL Interaction Centre, University College London, London, UK
†School of Psychology, Queen's University Belfast, UK
Email: (harry.griffin, m.aung, ucabbro, ucjt511, n.berthouze)@ucl.ac.uk
Email: (G.McKeown, w.curran)@qub.ac.uk

***Abstract*—Despite the importance of laughter in social interactions it remains little studied in affective computing. Respiratory, auditory, and facial laughter signals have been investigated but laughter-related body movements have received almost no attention. The aim of this study is twofold: first an investigation into observers' perception of laughter states (hilarious, social, awkward, fake, and non-laughter) based on body movements alone, through their categorization of avatars animated with natural and acted motion capture data. Significant differences in torso and limb movements were found between animations perceived as containing laughter and those perceived as non-laughter. Hilarious laughter also differed from social laughter in the amount of bending of the spine, the amount of shoulder rotation and the amount of hand movement. The body movement features indicative of laughter differed between sitting and standing avatar postures. Based on the positive findings in this perceptual study, the second aim is to investigate the possibility of automatically predicting the distributions of observer's ratings for the laughter states. The findings show that the automated laughter recognition rates approach human rating levels, with the Random Forest method yielding the best performance.**

Keywords: laughter, body movement, automatic emotion recognition, automatic laughter type recognition, laughter type perception

I. Introduction

The increasing use of virtual agents and robots in entertainment, collaborative, and support roles places ever greater demands on their ability to detect users' emotional state from various modalities (body movements, facial expressions, speech) and produce emotional displays. This is particularly true in socially complex human-computer interactions such as education, rehabilitation and health scenarios. In these situations emotionally expressive agents are much preferred by users [1].

Laughter is a ubiquitous and complex signal that remains relatively uninvestigated, in contrast to studies on other emotional expressions such as smiling [2]. Due to the range of vocal and physical expressions of laughter, its detection and synthesis are very challenging. Laughter does more than express hilarity. It can convey negative and mixed emotions and act as an invitation to shared expression [3]. At least 23 types of laughter have been identified (hilarious, anxious, embarrassed, etc.) [4] with each laughter type having its own social function. Hence, the ability to produce the appropriate type and intensity of laughter in response to a user's emotional signals, including laughter, would be a dramatic step forward in the realism and possibly efficacy of virtual agents.

There have been few studies on synthesizing laughter in virtual agents, most of which have focused on acoustics and the face [5], [6]. Urbain et al. present a laughter machine that is able to recognize laughter from sounds and give a response [7]. The distinctive respiration patterns of laughter have been widely corroborated [8] and integrated into anatomically inspired models of laughter [9]. Recently, Niewiadomski and Pelachaud investigated the coordination of virtual agents' laughter respiration behaviour with other visual cues; however, this work is mainly based on hilarious laughter [10]. A further difficulty for synthesis of laughter-related body movements is that stereotypical laughter actions, e.g. clutching ones abdomen, rocking back and forth, slapping one's leg, are well known but may be seen as exaggerated and unnatural.

Work on automatic recognition of laughter has also started to emerge but, as with the synthesis of laughter, has mostly focused the acoustic modality (e.g., [11]–[13]) and more recently on the combination of face and voice cues [14]. Less attention has been given to body laughter expressions. Whole-body postural changes and peripheral gestures associated with different types of laughter remain unelucidated. In [15], the authors use electromyographic sensors to measure diaphragmatic activity to detect laughter in people watching television. This is used to trigger laughter in nearby robotic dolls with the aim of enhancing the user's laughter.

More recently, there has been interest in creating automatic classifiers able to differentiate laughter types. To this end, motion descriptors based on energy estimates, correlation of shoulder movements and periodicity to characterise laughter have been investigated [16]. Using a combination of these measures a Body Laughter Index (BLI) was calculated. The BLIs of 8 laughter clips were compared with 8 observers' ratings of the energy of the shoulder movement. A correlation, albeit weak, between the observers' ratings and BLIs was found.

There has been growing evidence supporting the possibility of automatically discriminating between different emotions from various modalities: acoustics [17], facial expressions [18] and body [19]–[23]. Moreover, the study in [24] went further in trying to characterize different types of laughter. They investigated automatic discrimination of five types of acted laughter: happiness, giddiness, excitement, embarrassment and

hurtful. Actors were asked to enact these five emotions using both vocal and facial expressions whilst they were video-recorded. The video clips were labelled by expert observers who were also made aware of the intention of the actors. The results showed that automatic recognition based only on the vocal features reach higher accuracy (70% correct recognition) than when using both facial and vocal features (60% correct recognition) or facial features alone (40% correct recognition). While, on the basis of these results, the authors argue that vocal expressions carry more emotional information than facial expressions, it should be noted that the actors were asked to try to keep the head as still as possible so that it was always frontal to the video camera. These may have constrained and limited the way people expressed their laughter through their faces and head movements. In addition, the fact that the expressions were acted also raises the questions of how naturalistic they were. One could argue that we are better at acting an expression through our voice since we can hear it, while we cannot see our face. This is particularly true when the actors are not professionals but lay people.

In this study we investigate perception of laughter type from body movements and lay the groundwork for laughter type recognition from these cues. This study makes two contributions: first, by identifying body movements that are perceived as indicative of different types of natural laughter, it informs more convincing animation of laughter in avatars, which will increase their perceived conversational authenticity and emotional range. Second, it investigates if it is possible to automatically discriminate between different types of laughter by comparing a wide range of automated recognition methods.

II. Motion Data Collection

Users' perception of laughter-related body movements was investigated in a forced-choice perceptual experiment. Body movements captured during different types of natural and acted laughter were used to animate an avatar. Observers categorized the animations as hilarious, social, awkward, fake, or non-laughter. Naive observers' categorizations were used to allow analysis of the perception of body movements in the absence of other modalities e.g., verbal, facial, and in the absence of knowledge of the eliciting stimulus and context.

A. Laughter Capture

Nine pairs of participants took part in a motion capture recording session. The movements of one member of each pair (subjects - 3 male, 6 female, mean age 25.7) were captured using a whole-body inertial motion capture suit (Animazoo IGS-190). The suit was modified to maximize the sensitivity to spine and shoulder movements. Tasks to elicit laughter in both standing and sitting postures included word games, collaborative games (Pictionary) and humorous videos [25]. Laughter also occurred during conversation during "rest" periods. The subjects also produced fake laughter on request.

B. Stimulus Preparation

Using video recordings of the motion capture session, we segmented laughter episodes and gave them preliminary labels: hilarious; social (back-channeling, polite, conversational laughter); awkward (involving a negative emotion such as embarrassment or discomfort on another's behalf); or fake. In total, 508 laughter segments and 41 randomly located non-laughter segments, some containing other behaviour such as talking, were identified. The motion capture data from these segments were used to animate an avatar defined by the positional co-ordinate triplets of 26 anatomical points over the whole body. The anatomical proportions were the same for all animations (Figure 1). Viewing angle was standardized to a slightly elevated ¾ viewpoint, although models were free to walk and turn in the standing tasks. One hundred and twenty-six animations (experimenter labels: 34 hilarious, 43 social, 16 awkward, 19 fake, 14 non-laughter - mean duration = 4.1s, SD = 1.8s) were selected as stimuli for the perceptual phase (non-laughter animations were chosen randomly from previous sample, with durations within the range of durations of laughter animations). This ratio of laughter types according to experimenter-determined labels was designed to match the frequency of laughter-types in a naturalistic database [4]. Note that the level of agreement between the experimenter-determined labels and observers' categorization is not of interest here; rather we wished to establish which body movements are perceived by the observers as indicative of different laughter types. Therefore this distribution of stimuli by experimenter-determined labels was implemented only with the aim of producing sufficient segments in each observer-determined category to allow valid statistical analysis of body movement. The observers' categorisations act as our ground truth and the experimenter determined labels are not used in the analysis.

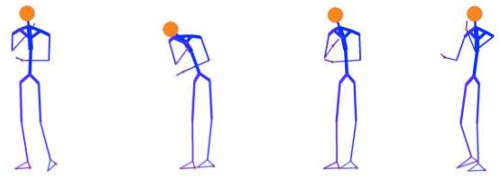

Fig. 1. Examples stills from the animated avatars

III. Perceptual Study:

A. Body Feature Analysis

Thirty-two observers (17 male, 15 female, mean age 33.0) viewed the clips of the animated avatar in random order and categorized each clip as hilarious, social, awkward, fake or non-laughter. No audio was presented with the animations.

The modal laughter category selected by the observers acted as the ground truth for the statistical analysis of body movement features [19]. The number of potential movement features that can be analyzed is large and increases exponentially if the interactions of multiple features are considered in combination. Therefore, our selection of features was based on previous findings in the literature [9], [26] and observers' comments in post-experiment interviews on which features they found useful in categorizing laughter. These included postural changes such as bending of the spine and gestures such as moving a hand toward the face or abdomen (Table I). Feature analysis was based on the position coordinate triplets of the relevant anatomical nodes. Maximum and minimum bending were calculated as greatest and smallest deviation respectively

TABLE I. LIST OF KNOWLEDGE BASED FEATURES TO BE ANALYSED.

Hands/gesture
Maximum, minimum and range of distance between hands
Maximum, minimum and range of distance of left hand from hip
Maximum, minimum and range of distance of right hand from hip
Maximum, minimum and range of distance of left hand from head
Maximum, minimum and range of distance of right hand from head
Shoulder movement
Correlation of left and right shoulder-hip distances
Range of azimuthal shoulder rotation
Spine and neck bending
Maximum, minimum and range of upper back bending
Maximum, minimum and range of lower back bending
Maximum, minimum and range of neck bending
Maximum, minimum and range of compound spine bending

from collinearity of the spine sections adjacent to the node in question. Range of bending was calculated as maximum bending minus minimum bending. Bending was calculated at each spine node including the neck, and collectively across all spine nodes (*compound bending*), defined as the sum of deviation from collinearity of all spine sections. Distances were calculated as Euclidean distances in 3D space. The features for hilarious, social and non-laughter segments were entered into separate one-way ANOVAs for standing and sitting segments (the independent variable was the modal observer categorization). Planned comparisons tested differences between laughter and non-laughter (hilarious and social vs. non-laughter) and between laughter types (hilarious vs. social).

B. Ground Truth from Observer Categorization

The mean number of observers who selected the modal category was 13.8 ($SD = 4.3$) with a maximum agreement of 29 of the 32 observers. Segments tied for the modal category were excluded from the body movement analysis, as were segments for which the modal category was selected by less than $\frac{1}{3}$ of observers ($< 11/32$). For all experimenter defined labels, the most common observer categorization was social or non-laughter. Too few awkward ($N = 4$) and fake ($N = 1$) remained so these were excluded from further analysis. Ninety-one segments (52 standing; 39 sitting) were entered into the final analysis of body movements.

C. Body Movements

For sitting laughter, ANOVAs revealed main effects of observer categorization on the range of distance between the hands, and the range of both hands' distance from the head and hip (all $F(2, 36) > 7$, $p \leq .003$); the range of azimuthal shoulder rotation ($F(2, 36) = 10.04$, $p < .001$); the range of bending at all spine and neck modes and of compound spine bending (all $F(2, 36) > 11$, $p < .001$); and the minimum bending at the upper back and neck (both $F(2, 36) > 4.5$, $p < .02$).

For all of these features, planned contrasts revealed significantly greater activity in laughter than non-laughter segments (all $t_{abs} > 2.5$, $p < .02$). Planned comparisons also revealed greater range of distances of both hands from the hip and of the left hand from the head in hilarious than social laughter (all $t_{abs} > 2$, $p < .04$); and a greater range of azimuthal shoulder rotation, greater range of bending at all spine and neck nodes and a greater range of compound spine bending in hilarious than social laughter (all $t_{abs} > 2$, $p < .05$).

For standing laughter, ANOVAs revealed main effects of observer categorization on the range of distance between the hands, the range of both hands' distance from the head and hip, the maximum distance of both hands from the hip and the minimum distance of the right hand from the head (all $F(2, 49) > 3$, $p < .05$); the range of bending and the maximum bending of upper and lower back and compound spine bending (all $F(2, 49) > 3$, $p < .05$).

Planned comparisons of these effects revealed greater range of hand-to-hand, hand-to-head, and hand-to-hip distances for both hands in laughter than non-laughter segments, and the range of right-hand-to-hip distances was greater in hilarious than social laughs (all $t_{abs} > 2.5$, $p < .02$); both hands moved further from the hip and the right hand moved closer to the head in laughter than non-laughter segments (all $t_{abs} > 3$, $p < .05$); the range of upper, lower and compound spine bending was greater for laughter than non-laughter segments and the range of upper and compound spine bending was greater for hilarious than social laughs (all $t_{abs} > 2$, $p < .05$), in addition the maximum compound spine bending was greater in laughter than non-laughter segments ($t_{abs} > 2.46$, $p = .018$).

IV. AUTOMATIC RECOGNITION

The second aim in this study is to investigate the possibility of automatically predicting the distributions of observers' ratings for the five types of laughter. The relative performances of a broad range of supervised machine learning methods are tested. In this part of the study we consider the distributions of the ratings from all 32 observers. This leads to a 5-output regression problem. If the frequencies of these ratings are normalised the values can be viewed as a degree of belief for each outcome and we also preserve a measure of observer agreement for each instance. This also removes the need to equate the most frequent label as a ground truth which is a weak assumption for instances with low agreement. Moreover, this will also allow for the full set of 126 instances to be used. The knowledge based features listed in Table I serves as part of the full feature set for recognition. We also include kinematically derived motion quantities analogous to the amount of energy expended. It has been shown that kinetic energy measures can contribute to the detection of laughter [16]. For three dimensional motion data a measure analogous to kinetic energy can be compactly calculated using the sum of the angular velocity at each joint over for each laughter segment [22]. Therefore, in the full feature set we also include the energy from five upper body articulations: left and right elbows, left and right shoulders and neck. Initial experiments showed a low degree of variance in lower body joints for this dataset and were therefore excluded.

A. Supervised Learning Models

Formally the problem consists of a set of $T = 5$ supervised regression tasks, one for each type of laughter (including 'non-laughter'). We denote by $x^i \in \mathbb{R}^d$, the vector of attributes describing instance i. We define the matrix of all of the training instances as $X \in \mathbb{R}^{d \times m}$, where m is the number of training instances and d being the dimensionality of the data. A distinct label y_t^i is provided for each task $t \in \{1 \dots T\}$, for instance i, taken from the frequency of observations. We denote $Y_t \in$

$\mathbb{R}^m$ as the vector label t for all instances. We also denote the corresponding model predicted output as $\hat{y}_t^i$.

a) k-Nearest Neighbour ($k-$NN): This is a simple model which assigns the value of the predicted output based on the K nearest training instances in the data space. We attain the necessary multiple outcome vector by using the means of the labels from the K nearest neighbours $N_K(x) \subset \{1, 2, \ldots, m\}$ of a given instance x. For a test instance x, the prediction is calculated by $\hat{y}_t^i = \frac{1}{K} \sum_{i \in N_K(x)} y_t^i$.

b) Multi Layer Perceptron with Softmax (MLP): The MLP is a widely used feed forward neural network that can be naturally applied to learn multiple regression tasks. For our purposes we further constrain the sum of the network outputs to 1 by using the softmax activation function [27]. This is an extension of the logistic function given by:

$$\hat{y}_t^i = \frac{\exp\left(q_t^i\right)}{\sum_{s=1}^{T} \exp\left(q_s^i\right)},$$

where q_t^i is the activation value for the output node for task t and input i.

c) Random Forest (RF): We also investigate the use of the Random Forest algorithm [28] to generate an ensemble of decision trees, using the mean of the ensemble as the final outcome. Each of these trees only has access to a set of δ attributes, randomly chosen when the tree was created. In the experiments conducted here, we have set the number of trees to 500, and the number of attributes considered for each tree $\delta = \left\lfloor \sqrt{d} \right\rfloor = 5$, as suggested in [29].

d) Linear and Kernel Ridge Regression (RR, KRR): This is a baseline regression approach. In the linear form, RR is based in solving the optimization problem $\min_{w_t} \left\| X^\top w_t - Y_t \right\|_2^2 + \lambda \left\| w_t \right\|_2$, where w_t represents the weight vector of the linear model $f_t(x) = \langle w_t, x \rangle$, $x, w_t \in \mathbb{R}^d$, for task $t \in \{1 \ldots T\}$. For convenience we denote as $\|\cdot\|_2$ the ℓ_2-norm of a vector. One can extend this approach to non-linear models by applying the kernel trick. In this case we have chosen the Gaussian kernel $\mathcal{K}(x,t) = \exp\left(\frac{-1}{\sigma^2} \left\| x - t \right\|_2^2\right)$.

e) Linear and Kernel Support Vector Regression (SVR, KSVR): Finally we implement Support Vector Regression to predict the degree of belief of each of the laughter type based on the frequency of the ratings for each instance. In the linear form, SVR is based on the optimization of the following problem:

$$\min_{w_t, \xi} \tfrac{1}{2} \left\| w_t \right\|^2 + C \sum_{i=1}^{m} \xi^i$$
$$\text{s.t} \begin{cases} \left| y_t^i - w_t^\top x^i \right| & \leq \varepsilon + \xi^i \\ \xi^i & \geq 0 \end{cases}$$

In that, $\varepsilon \geq 0$ is the deviation allowed from the ground truth labels y_t^i. This constraint is weakened in some points by adding an extra margin ξ^i. The degree of deviations larger than ε are adjusted by the second parameter $C \geq 0$. Similar to KRR, a non linear variant KSVR is also used in the comparison, employing also the Gaussian kernel.

B. Evaluation Metrics

In order to robustly evaluate the multiple outcomes of the models against the distribution of the observers categorisations, as suggested in [23], we apply four well established multi-score metrics over a number of instances M:

1) Mean Square Error: this is the standard loss function which is computed as:

$$MSE := \frac{1}{MT} \sum_{i=1}^{M} \sum_{t=1}^{T} \left(y_t^i - \hat{y}_t^i \right)^2$$

2) Cosine Similarity: finds the cosine of the angle between two vectors resulting in a maximum of 1 when the vectors are fully aligned.

$$CS := \frac{1}{M} \sum_{i=1}^{M} \frac{y^{i\top} \hat{y}^i}{\left\| y^i \right\|_2 \left\| \hat{y}^i \right\|_2}$$

3) Top Match Rate: evaluates the number of times the predicted top ranked label is the same as the top ranked label for the ground truth.

$$TMR := \frac{1}{M} \sum_{i=1}^{M} 1_{\left\{ \underset{1 \leq t \leq T}{\operatorname{argmax}}\, y_t^i = \underset{1 \leq t \leq T}{\operatorname{argmax}}\, \hat{y}_t^i \right\}}$$

where 1_A is a function on condition A.

$$1_A = \begin{cases} 1, & A\ is\ true \\ 0, & A\ is\ false \end{cases}.$$

4) Ranking Loss: this metric calculates the average fraction of label pairs that are reversely ordered for an instance. By ordering the label outcomes as:$\left(y_{l_1}^i \geq y_{l_2}^i \geq \ldots \geq y_{l_T}^i\right)$ the ranking loss predicted outputs can be calculated by:

$$RL := \frac{1}{M} \sum_{i=1}^{M} \frac{\sum_{j=1}^{T} \sum_{k=j+1}^{T} 1_{\left\{ \hat{y}_{l_j}^i < \hat{y}_{l_k}^i \right\}}}{T \times (T-1)/2}$$

where 1_A is the same function on condition A as for TMR.

C. Recognition Results

We implement and evaluate all of the models outlined above using a leave one subject out (LOSO) validation approach. This ensures instances from the same subject are not present in training, validation and test sets at the same time. We split the subjects into three groups: n training subjects, 1 validation subject to tune model parameters and 1 testing subject to assess performance. For each model this procedure is repeated 72 times (9 test subjects $\times 8$ validation subjects, accounting for all combinations) and the average results are reported. Parameter values were tuned over a set range for each of the models, the appropriate ranges were determined in initial experiments. The parameters adjusted are as follows: for $k-$NN:k; RR: λ; SVR: C; KSVR: C, σ; KRR: λ, σ; and MLP: n_{hidden} (the number of hidden layer nodes).

Table II compares the performances of all of the models using the four multi-score metrics. The results show mean (and

TABLE II. COMPARISON OF RECOGNITION PERFORMANCES. ↑ INDICATES HIGHER VALUES CORRESPOND TO BETTER PERFORMANCE AND ↓ INDICATES THE OPPOSITE. THE FIRST SEVEN ROWS CORRESPOND TO THE AUTOMATIC RECOGNITION MODELS, THE LAST ROW (IR) INDICATES THE MEAN LEVEL OF AGREEMENT BETWEEN OBSERVER GROUPS.

	MSE ↓	CS ↑	TMR ↑	RL ↓
k-NN	0.0151 (0.0041)	0.8825 (0.0300)	0.5272 (0.1658)	0.2998 (0.0517)
RR	0.0142 (0.0030)	0.8892 (0.0242)	0.4935 (0.2175)	0.2942 (0.0800)
KRR	0.0145 (0.0037)	0.8871 (0.0287)	0.5054 (0.2026)	0.2972 (0.0700)
SVR	0.0148 (0.0040)	0.8837 (0.0350)	0.4967 (0.2070)	0.3005 (0.0879)
KSVR	0.0149 (0.0039)	0.8842 (0.0302)	0.4815 (0.1965)	0.2989 (0.0791)
MLP	0.0192 (0.0066)	0.8536 (0.0450)	0.4837 (0.2112)	0.3195 (0.0668)
RF	**0.0101** (0.0036)	**0.9205** (0.0250)	**0.6620** (0.1665)	**0.2648** (0.0467)
IR	0.0217 (0.0032)	0.9457 (0.0081)	0.8489 (0.0291)	0.1003 (0.0092)

standard deviation) of each measure after the 72 runs. In order to understand how informative the form features alone (Table I) would perform, we also tested the models when trained without using the five energy based features. The results showed similar but reduced performances in comparison to the ones reported in Table II. For example the best performing scores without energy features were for the RF model with MSE: 0.0106, CS: 0.9163, TMR: 0.662, RL: 0.2712. This demonstrates the discriminatory power of the form features between laughter types. This supports previous results showing the importance of form in affective body expression recognition [30]. In addition, we also seek to understand the level of agreement between human observer groups. This calculation would provide a quantitative context when assessing the rates given in Table II. Using a simplified version of the approach proposed in [20], the raters were split randomly into two groups of 16 and the collective predictions of each group were computed. The same four measures used for evaluating the systems were applied to measure the agreement between these two predictions. We repeated this process 10000 times and computed the averages (and standard deviation). The results are reported in the last row of Table II as IR. We can see that the results obtained for the models are very similar to the inter-rater agreement measures for MSE and CS but are lower for TMR and RL.

Table III shows the F1-score and accuracy of the classifications for each laughter type from each of the models by assuming the most frequent observer label as the ground truth and the highest model output as the prediction. This can be viewed as treating the data as a classification problem. Within the 126 instances there were only 6 instances where 'awkward' was the most frequent label and 5 instances for 'fake', whereas the number of instances for 'hilarious', 'social', and 'non-laughter' were 25, 46, and 44 respectively. Moreover, for some of the subjects these classes do not occur if ground truth is considered in this way. Since we use LOSO classification performance can not be measured, therefore we show the F1 and accuracy scores for the remaining classes in Table III.

V. DISCUSSION AND CONCLUSION

In this section we discuss the findings from the perceptual study and the investigation into automated recognition.

TABLE III. F1-SCORE (TOP) AND ACCURACY (BOTTOM) FOR EACH MODEL BASED ON THE MOST FREQUENT OBSERVER LABELS FOR THE THREE CATEGORIES WITH A SIGNIFICANT NUMBER OF INSTANCES.

	Hilarious	Social	Not a Laugh
k-NN	**0.5941** 0.6000	0.3864 0.3397	0.5498 0.6818
RR	0.5253 0.5700	0.2287 0.1712	0.5875 0.7869
KRR	0.5268 0.5900	0.2744 0.2174	0.6068 0.7585
SVR	0.5103 0.5600	0.2555 0.1902	0.5864 0.7813
KSVR	0.4840 0.4900	0.2894 0.2418	0.5676 0.7273
MLP	0.4175 0.4050	0.3797 0.3261	0.5995 0.6932
RF	0.5636 **0.6200**	**0.5562** **0.5516**	**0.7441** **0.8011**

Analysis based on observer categorization of avatar animations revealed diagnostic body movement features for laughter perception. The importance of spine movements in sitting and standing postures may reflect observers' sensitivity to the respiratory movements that generate characteristic laughter vocalizations and cause the spine to bend [9]. Similarly, that hilarious laughter had a greater range of spine bending than social laughter may be due to the energetic nature of hilarious laughter relative to more controlled social laughter.

The range of azimuthal shoulder rotation was greater in laughter than non-laughter in the sitting but not standing posture. When standing, models were free to turn, whereas in the more constrained sitting condition shoulder rotation may have been indicative of an energetic laughter episode. Alongside the findings on spine bending, this hints that greater upper body movement may indicate laughter. It is counter-intuitive that any large upper body movement indicates laughter, so observers' perception of laughter compared to energetic, non-laughter movements, e.g. coughs, should be investigated.

The range of distance between the hands was greater in laughter than non-laughter segments, also indicating discrimination based on the overall amount of movement. An alternative explanation is the presence of specific gestures such as pointing to laughter-eliciting stimuli. Standing laughter segments had a smaller minimum right hand to head distance than those categorized as non-laughter, suggesting that moving the hand near or onto the face was seen as indicative of laughter. This is of particular interest, since this gesture is incidental to the core process of laughing; however, the timing of this gesture may be crucial in conveying the presence and nature of laughter and such temporal factors merit further investigation. For example the study reported in [31] shows that local temporal dynamics improves the automatic discrimination between affective body expressions.

There was insufficient consensus on awkward and fake laughter to draw conclusions on body movements indicative of these laughter types. These laughter types may be too emotionally and socially complex, or too infrequent in real life, for observers to have a clear mental model of the body movements associated with them. Alternatively these types of laughter may be indistinguishable, on the basis of body movements alone, from hilarious or social laughter, or from non-laughter speech. Further information, such as vocalizations, facial expressions, and context may be necessary for observers to disambiguate

them.

Although we optimized capture of shoulder and spine movement, the avatar animations were unable to show non-rigid deformation of the avatar sections (shoulder movement was shown through relative movements of rigid sections). Non-rigid deformations of the torso from respiratory action may be important in animating naturalistic laughter [9]. In addition our equipment did not capture hand gestures so the precise nature of arm and hand movements may have been ambiguous to observers, for example, they may have been unable to distinguish a pointing gesture from a palm-up gesture. Annotation of the video recordings of these sessions in future will identify meaningful gestures and, when these can be animated, allow us to analyse their contribution towards the perception of different laughter types.

Ultimately the capture of body movements using more accessible technology e.g., Microsoft Kinect, will make laughter detection ubiquitous in interactive systems. Our findings suggest that torso bending movements, possibly driven by respiratory actions, and peripheral gestures are used by observers to detect and classify laughter, and that these should be included when animating laughter. The resting posture, e.g. sitting vs. standing, should also be considered as it affects laughter diagnostic movements, e.g, shoulder rotation. Future work should cover more complex laughter, e.g. awkward, that we were unable to reliably elicit in this study. The sex, age, cultural background and personality of the laughter and observer should also be further considered, for example, laughter produced by extroverts and introverts may vary and specific attitudes towards laughter may affect the perception of the emotional content of the laughter. Some of these factors have been investigated in [32] using the same set of body laughter stimuli used in our study. The role of body movements may be more complex in multimodal displays than in this uni-modal study and our findings should be validated with simultaneous facial and audio information to establish their applicability in functional human-avatar interactions. The temporal dependencies of laughter signals between these modalities and within the body-movement channel will need to be carefully considered in these scenarios as the perceived emotional content of laughter may be strongly dependent on the order, duration and temporal profile, e.g. onset and offset speed, of these signals.

The results on automatic recognition (Tables II and III) demonstrate the effectiveness of the non parametric model RF. The relative poorer performance of the parametric models could be partially explained by the LOSO validation process used to tune the model parameters. Recalling that LOSO separates the training, validation and test sets by subject, this shows that they may have been prone to idiosyncratic effects during this tuning; this did not effect the RF model as no pre-tuning was done. In contrast, the other models showed a significant dependency on their respective tuned parameters k, λ, C, σ and n_{hidden}. The processing times for all of models are similar and are within the same order of magnitude with the exception of the MLP which required up to 10 times longer depending on n_{hidden}. When considering MSE and CS scores the recognition methods show a good performance. These metrics are more sensitive to the distribution of observer labels upon which all of the models are trained. It can be concluded that our full feature set used in this study is descriptive and appropriate for learning the observer distributions, with the worst performing model MLP still returning high scores. In contrast when considering TMR and LR the performances for all of the models return mediocre scores. However, in principle, this is not unexpected since all the methods are regression models by design.

Table III shows F1 classification scores for three categories: hilarious, social and non-laughter. The most readily classified category is non-laughter with social as the most difficult to discriminate. This shows the feature set used in this study could be salient for classifying non laughter from body movements. Nevertheless, they are still descriptive for the discrimination of the other classes well above chance level (20%). It is also worth noting that the MSE and CS rates for all of the models are similar to the MSE and CS scores for the inter observer group agreement. Though it must be noted that this is not directly comparable since the values in Table II stem from all 32 observers and the values calculated for IR stem from two groups of 16. Nevertheless, it does provide an indicator of the model performances relative to human recognition rates.

Future work should include the in-depth analysis of the decision tree ensembles within the RF model. This could give insight into the various features and corresponding thresholds that have the most discriminatory power and could further inform the design of improved recognition systems. Furthermore, methods to account for idiosyncratic artifacts should be considered such as individual bias removal [22] or transfer learning methods [33].

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 270780. We thank all those who participated in our experiments, Jianchuan Qi for his help in collecting the motion capture data and the members of the ILHAIRE consortium for their comments.

REFERENCES

[1] C. Creed and R. Beale, "User interactions with an affective nutritional coach," *Interacting with Computers*, vol. 24, no. 5, pp. 339–350, 2012.

[2] M. Ochs, R. Niewiadomski, P. Brunet, and C. Pelachaud, "Smiling virtual agent in social context," *International journal Cognitive Processing*, pp. 1–14, 2011.

[3] E. Holt, "The last laugh: Shared laughter and topic termination," *Journal of Pragmatics*, vol. 42, pp. 1513–1525, 2010.

[4] G. McKeown, R. Cowie, W. Curran, W. Ruch, and E. Douglas-Cowie, "Ilhaire laughter database," in *Proceedings of 4th International Workshop on Corpora for Research on Emotion, Sentiment & Social Signals, LREC*, 2012, pp. 32–35.

[5] D. Cosker and J. Edge, "Laughing, crying, sneezing and yawning: Automatic voice driven animation of non-speech articulations," in *Proceedings of Computer Animation and Social Agents, CASA*, 2009.

[6] R. Niewiadomski, J. Urbain, C. Pelachaud, and T. Dutoit, "Finding out the audio and visual features that influence the perception of laughter intensity and differ in inhalation and exhalation phases," in *Proceedings of 4th International Workshop on Corpora for Research on Emotion, Sentiment & Social Signals, LREC*, 2012.

[7] J. Urbain, R. Niewiadomski, E. Bevacqua, T. Dutoit, A. Moinet, C. Pelachaud, B. Picart, J. Tilmanne, and J. Wagner, "Avlaughtercycle. enabling a virtual agent to join in laughing with a conversational partner using a similarity-driven audiovisual laughter animation," *Journal of Multimodal User Interfaces*, vol. 4, pp. 47–58, 2010.

[8] M. Filippelli, R. Pellegrino, I. Iandelli, G. Misuri, J. Rodarte, R. Duranti, V. Brusasco, and G. Scano, "Respiratory dynamics during laughter," *Journal of Applied Physiology*, vol. 90, pp. 1441–1446, 2001.

[9] Z. V. DiLorenzo, P. and B. Sanders, "Laughing out loud: control for modeling anatomically inspired laughter using audio," *In ACM Transactions on Graphics*, vol. 27, p. 125, 2008.

[10] R. Niewiadomski and C. Pelachaud, "Towards multimodal expression of laughter," in *Intelligent Virtual Agents*. Springer, 2012, pp. 231–244.

[11] C.-H. Chou, C.-H. Li, B.-W. Chen, J.-F. Wang, and P.-C. Lin, "A real-time training-free laughter detection system based on novel syllable segmentation and correlation methods," in *Awareness Science and Technology (iCAST), 2012 4th International Conference on*. IEEE, 2012, pp. 294–297.

[12] K. Laskowski, "Contrasting emotion-bearing laughter types in multiparticipant vocal activity detection for meetings," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. IEEE, 2009, pp. 4765–4768.

[13] M. Miranda, J. A. Alonzo, J. Campita, S. Lucila, and M. Suarez, "Discovering emotions in filipino laughter using audio features," in *Human-Centric Computing (HumanCom), 2010 3rd International Conference on*. IEEE, 2010, pp. 1–6.

[14] S. Petridis and M. Pantic, "Audiovisual discrimination between laughter and speech," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 5117–5120.

[15] S. Fukushima, Y. Hashimoto, T. Nozawa, and H. Kajimoto, "Laugh enhancer using laugh track synchronized with the user's laugh motion," in *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '10. New York, NY, USA: ACM, 2010, pp. 3613–3618.

[16] M. Mancini, G. Varni, D. Glowinski, and G. Volpe, "Computing and evaluating the body laughter index," in *Human Behavior Understanding*. Springer, 2012, pp. 90–98.

[17] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572–587, 2011.

[18] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 1, pp. 39–58, 2009.

[19] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: a survey," *IEEE Trans. Affective Computing*, vol. 4, pp. 15–33, 2013.

[20] A. Kleinsmith, N. Bianchi-Berthouze, and A. Steed, "Automatic recognition of non-acted affective postures," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 41, no. 4, pp. 1027–1038, 2011.

[21] G. Castellano, S. D. Villalba, and A. Camurri, "Recognising human emotions from body movement and gesture dynamics," in *Affective computing and intelligent interaction*. Springer, 2007, pp. 71–82.

[22] D. Bernhardt and P. Robinson, "Detecting affect from non-stylised body motions," in *Affective Computing and Intelligent Interaction*. Springer, 2007, pp. 59–70.

[23] H. Meng, A. Kleinsmith, and N. Bianchi-Berthouze, "Multi-score learning for affect recognition: the case of body postures," in *Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 225–234.

[24] C. Galvan, D. Manangan, M. Sanchez, J. Wong, and J. Cu, "Audiovisual affect recognition in spontaneous filipino laughter," in *Knowledge and Systems Engineering (KSE), 2011 Third International Conference on*. IEEE, 2011, pp. 266–271.

[25] G. McKeown, W. Curran, C. McLoughlin, H. J. Griffin, and N. Bianchi-Berthouze, "Laughter induction techniques suitable for generating motion capture data of laughter associated body movements," in *2nd International Workshop on Emotion Representation, Analysis and Synthesis in Continuous Time and Space (EmoSPACE)*, 2013.

[26] W. Ruch and P. Ekman, "The expressive pattern of laughter," *Emotion, qualia, and consciousness*, pp. 426–443, 2001.

[27] J. S. Bridle, "Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition," in *Neurocomputing*. Springer, 1990, pp. 227–236.

[28] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[29] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, "Random forest: a classification and regression tool for compound classification and qsar modeling," *Journal of chemical information and computer sciences*, vol. 43, no. 6, pp. 1947–1958, 2003.

[30] A. Kleinsmith and N. Bianchi-Berthouze, "Form as a cue in the automatic recognition of non-acted affective body expressions," *Lecture Notes in Computer Science*, vol. 6874, pp. 155–164, 2011.

[31] A. Kleinsmith, T. Fushimi, and N. Bianchi-Berthouze, "An incremental and interactive affective posture recognition system," in *International Workshop on Adapting the Interaction Style to Affective Factors, in conjunction with the International Conference on User Modeling*, 2005.

[32] G. McKeown, W. Curran, D. Kane, R. McCahon, H. Griffin, C. McLoughlin, and N. Bianchi-Berthouze, "Human perception of laughter from context-free whole body motion dynamic stimuli," in *International Conference on Affective Computing and Intelligent Interaction*, 2013, in press.

[33] B. Romera-Paredes, M. Aung, M. Pontil, A. Williams, P. Watson, and N. Bianchi-Berthouze, "Transfer learning to account for idiosyncrasy in face and body expressions," *Automatic Face and Gesture Recognition*, 2013.